Errata for
"Identifying sequence-structure pairs undetected by sequence alignments"

Running title: Sequence - Structure Alignments

Key words: empirical potentials / inverse protein folding / protein fold recognition / sequence-structure alignment / threading and inverse threading with gaps and insertions

Sanzo Miyazawa[1] and Robert L. Jernigan[2]

[1]Faculty of Technology
Gunma University
Kiryu, Gunma 376
Japan
Phone: +81-277-30-1940
Fax: +81-277-40-1029
E-Mail: miyazawa@smlab.sci.gunma-u.ac.jp

[2]Room B-116, Bldg. 12B, MSC 5677
Laboratory of Experimental and Computational Biology
DBS, National Cancer Institute
National Institutes of Health
Bethesda, MD 20892-5677
U.S.A.

(May 7, 2009)

1. Equation 23 should have been printed as follows.

$$\begin{aligned}
\mathcal{P}(p,q) &= \frac{1}{\mathcal{Z}} \sum_{A \text{ with } (p,q)} \exp[-\beta\mathcal{E}(\{s_p\}|\{i_q\},A)] && (22)\\
&\simeq \frac{1}{\mathcal{Z}} \mathcal{Z}_{p-1,q-1}\ \exp[-\beta\mathcal{E}(\{s_p\}|i_q,\mathcal{P}(p',q'))]\ \mathcal{Z}'_{p+1,q+1} && (23)
\end{aligned}$$

2. In page 463,

   > **(i)** Set p_1 and p_2 to the N-terminal and C-terminal site position of a sequence segment to align, and q_1 and q_2 to the N-terminal and C-terminal site position of a partial structure to align.

   should be

   > **(i)** Set p_1 and p_2 to the N-terminal and C-terminal site position of a partial structure to align, and q_1 and q_2 to the N-terminal and C-terminal site position of a sequence segment to align.

3. In the section of “Datasets of protein structures” in page 465,

   > As a result, our set of superfamily representatives includes 308 proteins, the set of family representatives has 440 proteins and the set of domain representatives has 988 proteins.

   should be

   > As a result, our set of superfamily representatives includes 318 proteins, the set of family representatives has 440 proteins and the set of domain representatives has 988 proteins.

4. In Table IV,

   > false negative

   should be read as

   > false positives

   and vice versa.

5. In the section of References,

   Vendruscolo, M. and Domany, M. (1998)

   should be

   Vendruscolo, M. and Domany, E. (1998)

6. Figure 8a may be replaced by

```
minimum energy alignment
    sequence  3GRS 364    YNNIPTVV-FSHPPIGTVGLTEDEAIHKYGIENVKTYSTS FTPMYHAVTKRKTKCVM
    matched to:                             |  |  |  | |    |                 | |
    structure 1NPX 322    GVQGSSGLAVFDYKFASTGINEVMA-QKLGK-ETKAVTVV -EDYLMDFNPDKQKAWF
probability alignment
    sequence  3GRS 364    YNNIPTVV-FSHPPIGTVGLTEDEAIHKYGIENVKTYSTS FTPMYHAVTKRKTKCVM
    matched to:                      ?      |  |  ?  | | |  |   ?             | |
    structure 1NPX 322    GVQGSSGLAVFDYKFASTGINEVM-AQKLGKE-TKAVT-V VEDYLMDFNPDKQKAWF
                          7777664334334698999887541577776424333203 34444444455566666

              1NPX 322         bbbbb  bbbbbb  aaaa aaaa     bbbb b bbbb        bbbbb
                            ######                 ######                  #########
              3GRS 364         bbb     bbbbbbbaaaaaaaaaa  bbbbbbbb b aaaaa       bbb
minimum energy alignment
    structure 3GRS 364    YNNIPTVVFSHP PIGTVGLTEDEAIHKYGIENVKTYSTS FTPMYHAVTKRKTKCVM
    matched to:                                      | | |            |       | |
    sequence  1NPX 322    GVQGSSGLAVFD YKFASTGINEVMAQKLGKETKAVTVVE DYLMDF--NPDKQKAWF
probability alignment
    structure 3GRS 364  --YNNIPTVVFSHP-PIGTVGLTEDEAIHKYGIENVKTYSTS-FTPMYHAVTKRKTKCVM
    matched to:           ??          ?     |  |?    | | |       ?? ???       | |
    sequence  1NPX 322  GV--QGSSGLAVFDYKFASTGINE-VMAQKLGKETKAVTVVEDY---LMDFNPDKQKAWF
                        432233444444430345556554145667776543322222021112233566677777


minimum energy alignment
    sequence  3GRS 420  KMVCANKEEKVVGIHMQG-LGCDEMLQGFAVAVKMGATKADFDNT-VAIHPTSSEELVTL
    matched to:         | |         |                  |     |  |         |
    structure 1NPX 376  KLVYDPETTQILGAQLMSKADLTANINAISLAIQAKMTIEDLAYADFFFQPAFDKPWNII
probability alignment
    sequence  3GRS 420  KMVCANKEEKVVGIHM-QGLGCDEMLQGFAVAVKMGATKADFDNT-VAIHPTS-SEE-LV
    matched to:         | |         |   ?              |     |  |    ?    |      ?
    structure 1NPX 376  KLVYDPETTQILGAQLMSKADLTANINAISLAIQAKMTIEDLAYADFFFQPAFDKPWNII
                        666666677777765404567999999888888888888887643445543322222122

              1NPX 376  bbbb        bbbbbb       aaaaaaaaaa  aaaaaaa             aaa
                        #####              #########################
              3GRS 420  bbbbb b  bbbbbbbbbb      aaaaaaaaaa  aaaaaaa           aaaaa
minimum energy alignment
    structure 3GRS 420  KMVCA-NKEEKVVGIHMQGLGCDEMLQGFAVAVKMGATKADFDNT----VAIHPTSSEEL
    matched to:         | |             |              |     |  |
    sequence  1NPX 376  KLVYDPETTQILGAQLMSKADLTANINAISLAIQAKMTIEDLAYADFFFQPAFDKPWNII
probability alignment
    structure 3GRS 420  KMVCANKEEKVVG-IHMQGLGCDEMLQGFAVAVKMGATKADFDN----TVAIHPTSSEEL
    matched to:         | |         |?  |              |     |  |       ?     ??????
    sequence  1NPX 376  KLVYDPETTQILGAQLMSKADLTANINAISLAIQAKMTIEDLAYADFF------------
                        776653443444303444334444445555567778888887643356222222111100


minimum energy alignment                               min.ene. rmsd #aligned identities
    sequence  3GRS 478                  R-----------
    matched to:                                         -26.4    3.9  112      0.12
    structure 1NPX 436                  NTAALEAVKQER
probability alignment
    sequence  3GRS 476   TLR            ------------
    matched to:          ??                                      3.7  108      0.12
    structure 1NPX 436   ---            NTAALEAVKQER
                         011            246789999999             3.0   73

              1NPX 436                  aaaaaaaaaaa
                                            ########

              3GRS 475  aa
minimum energy
    structure 3GRS 475  VTLR                --------
    matched to:          |                              -20.0    4.3  113      0.11
    sequence  1NPX 436  NTAA                LEAVKQER
probability alignment
    structure 3GRS 475  VTLR------------------------
    matched to:         ???                                      3.5   92      0.12
    sequence  1NPX 424  ----FQPAFDKPWNIINTAALEAVKQER
                        0112533343233344344577788999             3.0   45
```